NEAR-OPTIMAL TIME-SPACE TRADEOFF
FOR ELEMENT DISTINCTNESS

Andrew Chi-Chih Yao

CS-TR-139-88

March 1988
Near-Optimal Time-Space Tradeoff for Element Distinctness

Andrew Chi-Chih Yao

Department of Computer Science
Princeton University
Princeton, New Jersey 08544

Abstract

It was conjectured in Borodin et al. [J. Comput. System Sci. 22 (1981), pp. 351-364] that to solve the element distinctness problem requires $TS = \Omega(n^2)$ on a comparison-based branching program using space $S$ and time $T$, which, if true, would be close to optimal since $TS = O(n^2 \log n)$ is achievable. Recently, Borodin et al. [SIAM J. on Comput. 16 (1987), pp. 97-99] showed that $TS = \Omega(n^{3/2}(\log n)^{1/2})$. In this paper, we will show a near-optimal tradeoff $TS = \Omega(n^{2-\varepsilon(n)})$, where $\varepsilon(n) = O(1/(\log n)^{1/2})$. 

1 Introduction

In Cobham's classic paper [C], time-space tradeoffs were established for one-tape Turing machines. In recent years, a number of time-space tradeoff results have been obtained for various computational models, such as Boolean and arithmetic circuits (Tompa [To]), a general sequential computing model (Borodin and Cook [BC]), multihead Turing machines (Duris and Galil [DG], Karchmer [K]), comparison-based branching programs (Borodin et al. [BFKLT], Yao [Y], Borodin et al. [BFMUW], Johnson [J], Karchmer [K]), and VLSI models (Thompson [Th], see Ullman [U] for a review of results). In this paper, we will establish a tradeoff result in the comparison-based branching program model, proving in weaker form an interesting conjecture of Borodin et al. [BFKLT].

Borodin et al. [BFKLT] proved a tradeoff $TS = \Omega(n^2)$ for sorting $n$ numbers on a comparison-based branching program, but were not able to establish a similar tradeoff for any decision problem in their model. They conjecture, however, that the tradeoff $TS = \Omega(n^2)$ is also true for the element distinctness problem. This, if proved, would be close to the best possible, since an upper bound $TS = O(n^3 \log n)$ is achievable for sorting, and hence for the element distinctness problem. Recently, Borodin et al. [BFMUW] gave a partial resolution to the above conjecture, showing in the same model that $TS = \Omega(n^{3/2} \log n)^{1/2})$. In this paper, we will prove that $TS = \Omega(n^{2-\epsilon(n)})$, where $\epsilon(n) = O(1/(\log n)^{1/2})$. As mentioned earlier, such a tradeoff is nearly the best possible.

Let $x_1, x_2, \ldots, x_n$ be $n$ elements chosen from a linearly ordered set $(D, \leq)$. The element distinctness problem (on $n$ elements) is to decide whether all $x_i$ are distinct. Following [BFMUW], a comparison branching program $A$ is a labeled directed acyclic graph with a distinguished nonsink node, called the source. Each nonsink node is labeled by a comparison $x_i : x_j$ with $i \neq j$, and has three outgoing edges, labeled by $<, =, >$, respectively. The sinks are labeled by either "accept" or "reject". An input $\bar{x} = (x_1, x_2, \ldots, x_n) \in D^n$ starts at the source and traverses $A$, making comparisons and branching according to the outcomes, until a sink is reached. The input is accepted if and only if it reaches a sink with an "accept" label. The capacity of $A$ is the base-2 logarithm of the number of nodes. The length of $A$, or the time $T$ used by $A$, is the length of the longest path starting with source. We say that $A$ is an algorithm for the element distinctness problem, if $\bar{x}$ is accepted when and only when all $x_i$ are distinct. Let $A_n$ denote the set of all algorithms for the element distinctness problem. We now state our main result.

Theorem 1 Any $A \in A_n$ with capacity $S$ and time $T$ must satisfy $TS = \Omega(n^{2-\epsilon(n)})$ for large $n$, where $\epsilon(n) = 5/(\ln n)^{1/2}$. 

2
2 Preliminaries

2.1 Overview

As discussed in [BFKLT], by a result of Nick Pippenger, we can assume without loss of
generality that $A$ is \textit{leveled}, i.e. each node is assigned a nonnegative integral level number, and
each edge goes from a node at level $i$ to a node at level $i + 1$; the source node is the only node
at level 0, and all sinks are at level $T$. From now on, all branching programs will mean leveled
comparison branching programs.

We review the ideas involved in the proof of $TS = \Omega(n^{3/2}(\log n)^{1/2})$ in [BFMUW]. (Some of
these ideas originated in [BFKLT].) Let $A \in \mathcal{A}_n$. For any input $\vec{x} = (x_1, x_2, \ldots, x_n)$ with distinct
$x_i$, the sequence of comparisons made by $A$ must include all the "adjacent" ones, i.e. comparisons
of the form $x_{ij} : x_{ij+1}$ if the input satisfies $x_{i_1} < x_{i_2} < \cdots < x_{i_n}$; otherwise we could have two
identical $x_i$. The idea is to show that any branching program of length less than or equal to
$n_0$, where $n_0 = (nS/(16e))^{1/2}$, can make more than $S$ adjacent comparisons only for a very small
fraction of the $n!$ possible linear orderings in the input. Thus, if we divide $A$ into consecutive
blocks of $n_0$ levels each, there must be at least $(n-1)/S$ such blocks in order to perform the
needed $n-1$ adjacent comparisons for all linear orderings. This proves $T \geq n_0(n-1)/S$, and
hence $TS = \Omega(n^{3/2}S^{1/2}) = \Omega(n^{3/2}(\log n)^{1/2})$ as $S = \Omega(\log n)$.

To prove Theorem 1, we will adopt the same general approach. We will show that any
branching program of length less than or equal to $n_1$, where $n_1 = n^{1-\epsilon(n)}$, can make more than
$S \cdot n^{\epsilon(n)}$ adjacent comparisons only for a very small fraction of the possible linear orderings. The
asserted tradeoff then follows the same line reasoning as before.

2.2 Terminology

Let $W = \{w_1, w_2, \ldots, w_n\}$ be any nonempty finite set. A \textit{linear ordering} on $W$ is a sequence
$\sigma = (w_{i_1} < w_{i_2} < \cdots < w_{i_n})$, in which each element of $W$ appears exactly once. Let $\Gamma(W)$ denote
the set of all linear orderings on $W$.

Let $P = (\prec_P, W)$ be a partial order on $W$. A linear ordering $\rho$ is said to be \textit{consistent with} $P$,
if $w \prec_P w'$ implies $w < w'$ in $\rho$. Let $\Delta(P)$ denote the set of all linear orderings on $W$ consistent
with $P$.

Suppose that $W' \subseteq W$ and $\sigma = (w_{r_1} < w_{r_2} < \cdots < w_{r_m}) \in \Gamma(W')$. A comparison $w_i < w_j$
is \textit{adjacent} in $\sigma$, if $w_i, w_j$ are adjacent in the linear ordering $\sigma$, i.e. there exists an $m$ such that
$i = r_m$ and $j = r_{m+1}$. For any $\rho \in \Gamma(W)$, let $\rho|_{W'}$ denote the $\sigma \in \Gamma(W')$ obtained from the
restriction of $\rho$ to $W'$. For example, if $\rho = (w_3 < w_2 < w_5 < w_1 < w_4)$ and $W' = \{w_2, w_4, w_5\}$,
then $\rho|_{W'} = (w_2 < w_5 < w_4)$. 
We will use the symbol $X$ to denote exclusively the set $\{x_1, x_2, \ldots, x_n\}$ of the $n$ input numbers. Let $C$ be a sequence of inequalities $(x_{i_1} < x_{j_1}, x_{i_2} < x_{j_2}, \ldots, x_{i_\ell} < x_{j_\ell})$. Its length $\ell$ is denoted by $|C|$, and $\text{support}(C)$ is the set of all $x_r$ involved in any comparisons in $C$, i.e., those $x_r$ with $r = i_s$ or $r = j_{s'}$ for some $s, s'$. We say that $C$ is nontrivial, if there is a linear ordering $(x_{r_1} < x_{r_2} < \cdots < x_{r_n})$ on $X$ such that all inequalities in $C$ are true. For any nontrivial $C$, its transitive closure defines a partial order $P_C$ on the set $\text{support}(C)$. Now, for each $C$ satisfying $2|C| \leq n$, let $V_C \subseteq X$ be a set disjoint from $\text{support}(C)$ with $|V_C| = 2|C| - |\text{support}(C)|$, and define $\text{support}'(C) = \text{support}(C) \cup V_C$. Let $P'_C$ be the partial order $P_C$ regarded as a partial order on $\text{support}'(C)$. In particular, if all the pairwise comparisons in $C$ are disjoint, then $V = \emptyset$ and $P'_C = P_C$.

Let $0 \leq r_1 < r_2 \leq \ell$ be two integers. For any $\sigma \in \Delta(P'_C)$, let $Z(C, r_1, r_2, \sigma)$ be the number of pairs $(i_s, j_{s+1}), r_1 < s \leq r_2$, such that the comparisons $x_{i_s} < x_{j_s}$ are adjacent in $\sigma$.

2.3 Main Lemma

Let $A$ be any branching program of variables $x_1, x_2, \ldots, x_n$. For any node $u$ and positive integer $s$, let $A[u, s]$ denote the sub-branching program of $A$ of length $s$ and rooted at node $u$. A path $\delta$ in $A$ is a sequence $u_1, e_1, u_2, e_2, \ldots, u_s, e_s$, where each $e_r$ is an edge from node $u_r$ to $u_{r+1}$; $s$ is the length of $\delta$. Let $C_{\delta}$ denote the sequence of comparison results ",$x_i < x_j\) and "$x_i = x_j\) obtained along the path $\delta$ ("$x_i > x_j\) will be written as "$x_j < x_i\). We are only interested in paths $\delta$ that (a) contain no edge labeled by "$=\)", and (b) have nontrivial $C_{\delta}$. From now on, when we speak of a path $\delta$, we require that the above two conditions be satisfied. Let $\Lambda_A$ be the set of all paths of length $T$ that begin with the source, where $T$ is the length of $A$. For any linear ordering $\rho \in \Gamma(X)$, let $\bar{x}_\rho$ denote an input $(x_1, x_2, \ldots, x_n)$ that satisfies all the inequalities in $\rho$. Let $\xi_{A, \rho} \in \Lambda_A$ be the path traversed by input $\bar{x}_\rho$.

Let $N_0 = 10^8$, $n \geq N_0$, $S > 0$, $t = [e^{(\ln n)^{k_1}}]$, and $k_0 = [\log_2(n/4)]$. Then $t, k_0 \geq 4$. For any integer $k > 0$, let $m_k = 2^{4k + 16}\epsilon S$, and $q_k = (4t)^{k - 10^8} S$.

Main Lemma Let $1 \leq k \leq k_0$. Suppose $A$ is a branching program of length $t^k$ and capacity $S$. Take a random $\rho$, uniformly chosen from $\Gamma(X)$, then $\Pr\{Z(C_{\delta}, 0, t^k, \rho|X') \geq m_k\} \leq q_k$, where $\delta = \xi_{A, \rho}$ and $X' = \text{support}'(C_{\delta})$.

Corollary Suppose that $A$ is a branching program of length $\leq t^{k_0}$ and capacity $S$. Take a random $\rho$, uniformly chosen from $\Gamma(X)$, and input $\bar{x}_\rho$ to $A$, then the probability that $A$ makes at least $m_k$ comparisons adjacent in $\rho$ is $\leq q_k$.

We remark that the Main Lemma is true for any choice of the sets $V_C$. However, the choice must be made before taking the random $\rho$. The corollary follows from the Main Lemma, since the introduction of additional elements $x_i$ into $\text{support}'(C_{\delta})$ will not increase the number of adjacent comparisons. It is this corollary that we will use later in the proof of Theorem 1. Before doing that,
we need to prove the Main Lemma. We first derive an auxiliary lemma in the next subsection. This will be used in Section 3 to prove the Main Lemma. The proof of Theorem 1 will then be given in Section 4.

2.4 An Auxiliary Lemma

Let $X' \subseteq X$ be nonempty, and $\sigma \in \Gamma(X')$. Suppose that $C$ is a nontrivial sequence of comparisons $x_i < x_j$ with $x_i, x_j \in X'$, with exactly $\ell$ of them being adjacent in $\sigma$. Let $A$ be a branching program of length $T$ with $n \geq 2T + |X'|$. For any path $\delta \in A_A$, let $W_\delta \subseteq X$ be such that $W_\delta$ is disjoint from $X' \cup \text{support}(C_\delta)$, and $|W_\delta| = 2T + |X'| - |X' \cup \text{support}(C_\delta)|$. Let $W_\delta' = W_\delta \cup X' \cup \text{support}(C_\delta)$. Clearly, $|W_\delta'| = 2T + |X'|$.

Now, take a random $\rho \in L(\sigma)$, uniformly chosen, and let $f(\sigma, C, A, m)$ be the probability that the number of comparisons of $C$ adjacent in $\rho|_{W_\delta'}$ is greater than or equal to $m$, where $\delta = \xi_{A, \rho}$.

**Lemma 1** Suppose $n \geq 2T + |X'|$. Then $f(\sigma, C, A, m) \leq \binom{\ell}{m} \cdot \left(\frac{|X'|}{2T}\right)^m$.

**Proof** Without loss of generality, assume that $X' = \{x_1, x_2, \ldots, x_a\}$, where $a = |X'|$. Let $C'$ be the set of the comparisons of $C$ that are adjacent in $\sigma$, say, $C' = \{x_{i_1} < x_{i_1+1}, x_{i_2} < x_{i_2+1}, \ldots, x_{i_\ell} < x_{i_\ell+1}\}$. We can also assume that $\ell > 0$ and $|X| > m > 0$; otherwise the lemma is trivially true.

We express $f$ in terms of a stochastic process. Take a random $\rho \in L(\sigma)$, and traverse the path $\xi_{A, \rho}$. Let us keep a sorted list $W'$; initially, $W'$ is the sorted version of $X'$. When we encounter a new node $u$ with a comparison $x_r : x_s$, we insert the elements in $\{x_r, x_s\} - W'$, if any, one at a time into the ordered list $W'$. Note that each new element, when added to an ordered list of $c$ elements, will be equally likely in any of the $c + 1$ ranks. When we reach the leaf, we add the $2T + |X'| - |W'|$ new elements of $W_\delta$, one at a time, into $W'$. Again, each new element is equally likely to occupy any of the ranks currently possible in $W'$. The quantity $f(\sigma, C, A, m)$ can thus be calculated as follows: We start with an ordered list of $|X'|$ items with $\ell$ of the intervals (between the $i_j$-th and the $i_{j+1}$-th items for $1 \leq j \leq \ell$) marked; then we sequentially insert new items into the list, each time the new item is equally likely to be inserted into any of the existing intervals; $f(\sigma, C, A, m)$ is the probability that, after $2T$ insertions, at least $m$ of the original $\ell$ marked intervals remain intact (no item has been inserted into these intervals).

We will obtain an upper bound on $f$. (Essentially, this is now reduced to a calculation which was done in [BFMUW].) Let us describe the above stochastic process using a sequence of $2T$ integers $j_1, j_2, \ldots, j_{2T}$, where $1 \leq j_r \leq |X'| + r$ is the rank of the $r$-th inserted item when it is being inserted. Thus, there are in all $\prod_{1 \leq r \leq 2T} (|X'| + r)$ configurations. To specify a configuration for which at least $m$ marked intervals remain intact, we first specify $m$ such intervals, and then specify the ranks of the inserted items by integers $j_1, j_2, \ldots, j_{2T}$, where $1 \leq j_r \leq |X'| + r - m$. 
The total number of such configurations is thus at most \((\ell \choose m) \frac{\prod_{1 \leq r \leq 2T}(|X'| + r - m)}{\prod_{1 \leq r \leq 2T}(|X'| + r)}\). It follows that

\[
f(\sigma, C, A, m) \leq \binom{\ell}{m} \frac{\prod_{1 \leq r \leq 2T}(|X'| + r - m)}{\prod_{1 \leq r \leq 2T}(|X'| + r)} \\
= \binom{\ell}{m} \frac{|X'|!}{(|X'| + 2T)!} \binom{|X'| - 1}{r - m + 1} \\
= \binom{\ell}{m} \frac{|X'| \cdot (|X'| - 1)(|X'| - m + 1)}{(|X'| + 2T)(|X'| + 2T - 1) \cdot \ldots \cdot (|X'| + 2T - m + 1)} \\
\leq \binom{\ell}{m} \left(\frac{|X'|}{2T}\right)^{m} \cdot \frac{1}{(2T)^{m}} \cdot \frac{1}{(2T)^{m}}.
\]

\[
\square
\]

3 Proof of the Main Lemma

We will prove the Main Lemma by induction on \(k \geq 1\). For \(k = 1\), we have \(m_{k} > t^{k}\). Since \(A\) can make only \(t^{k}\) comparisons, \(M_{A} = \emptyset\), and the Main Lemma is true in this case.

We now assume that \(1 < k \leq k_{0}\), and that the Main Lemma has been proved for all values less than \(k\).

Assume that the choice of \(V_{C}\) has been made. Let \(\rho \in \Gamma(X)\). Write \(\delta = \xi_{A, \rho}\) and \(X' = \text{support}'(C_{\delta})\). If \(Z(C_{\delta}, 0, t^{k}, \rho|_{X'}) \geq m_{k}\), then there is a \(1 \leq d_{\rho} \leq t\), such that \(Z(C_{\delta}, (d_{\rho} - 1)t^{k-1}, d_{\rho}t^{k-1}, \rho|_{X'}) \geq m_{k}/t\). Thus,

\[
\Pr\{Z(C_{\delta}, 0, t^{k}, \rho|_{X'}) \geq m_{k}\} \leq \sum_{1 \leq d \leq t} \Pr\{Z(C_{\delta}, (d - 1)t^{k-1}, dt^{k-1}, \rho|_{X'}) \geq m_{k}/t\}. \tag{1}
\]

We will show that, for each \(1 \leq d \leq t\),

\[
\Pr\{Z(C_{\delta}, (d - 1)t^{k-1}, dt^{k-1}, \rho|_{X'}) \geq m_{k}/t\} \leq 2q_{k-1}. \tag{2}
\]

This will complete the inductive proof of the Main Lemma, as it follows from (1) and (2) that

\[
\Pr\{Z(C_{\delta}, 0, t^{k}, \rho|_{X'}) \geq m_{k}\} \leq 2tq_{k-1} \leq q_{k}.
\]

Although we have made the choice of \(V_{C}\), which is needed to define the function \(Z\) above, we observe that the values of \(\Pr\{Z(C_{\delta}, 0, t^{k}, \rho|_{X'}) \geq m_{k}\}\) and \(\Pr\{Z(C_{\delta}, (d - 1)t^{k-1}, dt^{k-1}, \rho|_{X'}) \geq m_{k}/t\}\) are in fact independent of the choice of \(V_{C}\), as \(\rho\) is uniformly chosen from \(\Gamma(X)\). We will now evaluate \(\Pr\{Z(C_{\delta}, (d - 1)t^{k-1}, dt^{k-1}, \rho|_{X'}) \geq m_{k}/t\}\) with a special new choice of \(V_{C_{\delta}}\) to be described below.
Fix $1 \leq d \leq t$. Let $v_1, v_2, \ldots, v_r$ be the nodes of $A$ at level $(d - 1)t^{k-1}$. For each $i$, let $B_i$ be the set of paths $\beta$ of length $t^{k-1}$ starting at node $v_i$. For each $\beta \in B_i$, let us choose a subset $X_\beta \subseteq X$ such that (a) $X_\beta \cap \text{support}(C_\beta) = \emptyset$, and (b) $|X_\beta| + |\text{support}(C_\beta)| = 2t^{k-1}$. Let $X'_\beta = X_\beta \cup \text{support}(C_\beta)$, and $Q_\beta$ be the partial order on $X'_\beta$ generated by the inequalities in $C_\beta$. Let $\Psi_i = \bigcup_{\beta \in B_i} \Delta(Q_\beta)$.

For any $\sigma \in \Gamma(W)$, where $W \subseteq X$, let $L(\sigma)$ denote the set of all linear orderings $\rho \in \Gamma(X)$ that are consistent with $\sigma$.

**Fact 1** Let $\beta, \beta' \in B_i$. If $\sigma \in \Delta(Q_\beta) \cap \Delta(Q_{\beta'})$, then $\beta = \beta'$.

**Fact 2** The family $L(\sigma), \sigma \in \Psi_i$, form a partition of the set $\Gamma(X)$.

Fact 1 is true because any two distinct $\beta, \beta'$ must have a common node at which the comparison $x_r : x_s$ made gives opposite outcomes. We can thus write $\beta(i, \sigma)$ for the unique $\beta$ for which $\sigma \in \Delta(Q_\beta)$. To prove Fact 2, first we observe that every $\tilde{x}_\rho$ starting at $v_i$ will follow some path $\beta$. This shows that the union of $L(\sigma), \sigma \in \Psi_i$, contains $\Gamma(X)$. It remains to prove that, if $\sigma \neq \sigma'$, then $L(\sigma) \cap L(\sigma') = \emptyset$. This is clearly true when $\sigma, \sigma' \in \Delta(Q_\beta)$ for some common $\beta$. In the other case, $\sigma \in Q_\beta$ and $\sigma' \in Q_{\beta'}$ with $\beta \neq \beta'$. Any $\rho \in L(\sigma)$ and $\rho' \in L(\sigma')$ must be different, since $\tilde{x}_\rho, \tilde{x}_{\rho'}$ follow two different paths $\beta, \beta'$. This proves Fact 2.

For the discussion to follow, we will use the convention that a branching program of length 0 is a **null branching program**, denoted by $\Phi$. We agree that the expressions $M \Phi, \Phi M$ both stand for $M$, where $M$ is any branching program. A path of length 0 is the **null path**, denoted by $\psi$. A sequence of comparison inequalities is the **null sequence**, denoted by $\kappa$. We define $C_{\psi}$ to be $\kappa$.

Define $\text{support}(\kappa) = \emptyset$, and $\Lambda_\Phi = \{\psi\}$. For any $\rho \in \Gamma(X)$, let $\xi_{\Phi, \rho} = \psi$. The introduction of these notations is mainly for convenience, so as to avoid the necessity of discussing degenerate cases in the discussions to come.

Let $A' = A[\text{root},(d-1)t^{k-1}]$, and $A_\beta = A[u_\beta,t^k-t^{k-1}]$, where $u_\beta$ is the node reached by the last edge of the path $\beta$. Let $A'A_\beta$ denote the branching program one obtains by attaching a copy of $A_\beta$ to each leaf of $A'$. The length of $A'A_\beta$ is clearly $t^k - t^{k-1}$. We remark that, if $d = 1$, then $A' = \Phi$; if $d = t$, then $A_\beta = \Phi$ for all $\beta$.

Let $\eta = (i, \alpha, \beta, \gamma)$ be any quadruple, where $1 \leq i \leq r$, $\alpha \in \Lambda_{A'}$, $\beta \in B_i$, and $\gamma \in \Lambda_{A_\beta}$. Define $U_{\eta} = \text{support}(C_\alpha) \cup \text{support}(C_\gamma) \cup X_\beta \cup W_{\alpha, \gamma}$, where $W_{\alpha, \gamma} \subseteq X$ satisfies the conditions that it is disjoint from $\text{support}(C_\alpha) \cup \text{support}(C_\gamma) \cup X'_\beta$ and has cardinality $2t^k - |\text{support}(C_\alpha) \cup \text{support}(C_\gamma) \cup X'_\beta|$, but arbitrary otherwise. Thus, $|U_{\eta}| = 2t^k$. Note that the last edge of the path $\alpha$ does not have to end in the node $v_i$. For any $\rho \in \Gamma(X)$, define $Y(\rho, \eta) = 1$, if the number of comparisons of $C_\beta$ adjacent in $\rho|_{U_{\eta}}$ is at least $m_k/t$, and 0 otherwise.

Let $\delta \in \Lambda_A$. We will describe how to choose $V_{C_\delta}$. We can uniquely write $\delta = \alpha\beta\gamma$, where for some $i$, $\alpha$ is a path in $A'$, $\beta \in B_i$, and $\gamma$ is a path in $A_\beta$. Let $\eta(\rho) = (i, \alpha, \beta, \gamma)$. Define
\[ V_{C_\delta} = (X_{\beta} \cup W_{\alpha,\gamma}) - \text{support}(C_\alpha) - \text{support}(C_\gamma) \]. Then \( \text{support}'(C_\delta) = V_{C_\delta} \cup \text{support}(C_\delta) = X_{\beta}' \cup \text{support}(C_\alpha) \cup \text{support}(C_\gamma) \cup W_{\alpha,\gamma} \). Thus, \( \text{support}'(C_\delta) = U_{\eta}(\rho) \). It follows that \( Y(\rho, \eta(\rho)) = 1 \) if and only if \( Z(C_\delta, (d - 1)t^{k - 1}, dt^{k - 1}, \rho; X') \geq m_k / t \).

For \( 1 \leq i \leq r \), let \( R_i \) be the set of \( \rho \in \Gamma(X) \) such that input \( \tilde{x}_\rho \) will reach \( v_i \) in \( A \). Let \( \Psi_{i,1} \) be the set of \( \sigma \in \Psi_i \) such that \( C_{\beta(i,\sigma)} \) contains at least \( m_{k - 1} \) comparisons adjacent in \( \sigma \). Let \( \Psi_{i,2} \) be the set of \( \sigma \in \Psi_i \) such that \( |L(\sigma) \cap R_i| \leq |L(\sigma)|q_{k - 1}/(10 \cdot 2^S) \). Let \( \Psi_{i,0} = \Psi_i - \Psi_{i,1} - \Psi_{i,2} \).

Let \( p_i \) be the probability that \( \xi_{A,\rho} \) will reach \( v_i \) for a random \( \rho \), i.e. \( p_i = |R_i|/n! \). Let \( I \) be the set of \( i \) with \( p_i > 0 \). For any \( i \in I \) and \( \sigma \in \Psi_i \), let \( p_{i,\sigma} = |L(\sigma) \cap R_i|/|R_i| \), i.e. the probability that, given that \( v_i \) is reached by \( \xi_{A,\rho} \), \( \rho \) will be consistent with \( \sigma \). Let \( \Psi_i^t = \{ \sigma \mid p_{i,\sigma} > 0 \} \). Define \( \Psi_{i,j} = \Psi_{i,j} \cap \Psi_i^t \) for \( j \in \{0,1,2\} \). With the above choice of \( V_{C_\delta} \) for defining \( Z \), we have

\[
\Pr\{Z(C_\delta, (d - 1)t^{k - 1}, dt^{k - 1}, \rho; X') \geq m_k / t \} = \sum_{i \in I} \sum_{\sigma \in \Psi_i^t} p_i p_{i,\sigma} \sum_{\rho \in L(\sigma) \cap R_i} Y(\rho, \eta(\rho)) \}. \tag{3}
\]

We need three facts. Let \( i \in I \).

**Fact 3** \( \sum_{\sigma \in \Psi_{i,1}} p_{i,\sigma} \leq q_{k - 1} \).

**Fact 4** \( \sum_{\sigma \in \Psi_{i,2}} p_{i,\sigma} \leq q_{k - 1}/(10 \cdot 2^S) \).

**Fact 5** For each \( \sigma \in \Psi_{i,0} \) with \( L(\sigma) \cap R_i \neq \emptyset \),

\[
\frac{1}{|L(\sigma) \cap R_i|} \sum_{\rho \in L(\sigma) \cap R_i} Y(\rho, \eta(\rho)) \leq q_{k - 1}/10.
\]

If we apply the induction hypothesis to \( A[v_i, t^{k - 1}] \), we get Fact 3. We obtain Fact 4 from the following derivation, using Fact 2 in the last step,

\[
\sum_{\sigma \in \Psi_{i,2}} p_i p_{i,\sigma} = \sum_{\sigma \in \Psi_{i,2}} \frac{|R_i| |L(\sigma) \cap R_i|}{n! |R_i|} = \frac{1}{n!} \sum_{\sigma \in \Psi_{i,2}} |L(\sigma)| |L(\sigma) \cap R_i| \leq \frac{1}{10 \cdot 2^S} \sum_{\sigma \in \Psi_{i,2}} |L(\sigma)| \leq \frac{1}{10 \cdot 2^S} q_{k - 1}.
\]

8
We now prove Fact 5. Let \( i \in I \) and \( \sigma \in \Psi_{i,0} \). Take a random \( \rho \), uniformly chosen from \( L(\sigma) \). Write \( A'' = A_{\beta(i,\sigma)} \) and \( \zeta(\rho) = (i, \xi_{A'',\rho}, \beta(i, \sigma), \xi_{A'',\rho}) \). Then

\[
\frac{1}{|L(\sigma) \cap R_i|} \sum_{\rho \in L(\sigma) \cap R_i} Y(\rho, \eta(\rho)) = \frac{1}{|L(\sigma) \cap R_i|} \sum_{\rho \in L(\sigma) \cap R_i} Y(\rho, \zeta(\rho)) \\
\leq \frac{1}{|L(\sigma) \cap R_i|} \sum_{\rho \in L(\sigma)} Y(\rho, \zeta(\rho)) \\
= \frac{|L(\sigma)|}{|L(\sigma) \cap R_i|} \frac{1}{|L(\sigma)|} \sum_{\rho \in L(\sigma)} Y(\rho, \zeta(\rho)) \\
\leq \frac{10 \cdot 2^S}{q_{k-1}} \frac{1}{|L(\sigma)|} \sum_{\rho \in L(\sigma)} Y(\rho, \zeta(\rho))
\]  
(4)

It is clear that

\[
\frac{1}{|L(\sigma)|} \sum_{\rho \in L(\sigma)} Y(\rho, \zeta(\rho)) = f(\sigma, C_{\beta(i,\sigma)}, A' A_{\beta(i,\sigma)}, m_k/t).
\]  
(5)

Let \( \ell_\sigma \) denote the number of comparisons of \( C_{\beta(i,\sigma)} \) adjacent in \( \sigma \). Then \( \ell_\sigma \leq m_{k-1} \) as \( \sigma \in \Psi_{i,0} \). Using (5) and Lemma 1 (noting that \( n \geq 2t^k \)), we have

\[
\frac{1}{|L(\sigma)|} \sum_{\rho \in L(\sigma)} Y(\rho, \zeta(\rho)) \leq \left( \frac{\ell_\sigma}{m_k/t} \right) \left( \frac{2t^{k-1}}{2t^k - 1} \right)^{[m_k/t]} \\
\leq \left( \frac{m_{k-1}}{m_k/t} \right) (t-1)^{-[m_k/t]} \\
\leq \left( \frac{em_{k-1}}{(t-1)[m_k/t]} \right)^{[m_k/t]} \\
\leq \left( \frac{et}{10(t-1)} \right)^{m_k/t} \\
\leq \frac{1}{2^{1000S}}
\]  
(6)

It follows from (4) and (6) that

\[
\frac{1}{|L(\sigma) \cap R_i|} \sum_{\rho \in L(\sigma) \cap R_i} Y(\rho, \eta(\rho)) \leq \frac{10 \cdot 2^S}{q_{k-1}} \left( \frac{1}{2} \right)^{1000S} \\
\leq \frac{1}{10^{q_{k-1}}}
\]

This proves Fact 5.

We will now complete the proof of (2). From Facts 3, 4 and 5, we obtain from (3) that

\[
\Pr\{ Z(C_\delta, (d-1)t^{k-1}, dt^{k-1}, \rho | X') \geq m_k/t \}
\]

9
\[
= \sum_{i \in I} \sum_{\sigma \in \psi'_i} p_i p_{i,\sigma} \left\{ \frac{1}{|L(\sigma) \cap R_i|} \sum_{\rho \in L(\sigma) \cap R_i} Y(\rho, \eta(\rho)) \right\}
\leq \sum_{i \in I} \sum_{\sigma \in \psi'_{i,1}} p_i p_{i,\sigma} + \sum_{i \in I} \sum_{\sigma \in \psi'_{i,2}} p_i p_{i,\sigma}
+ \sum_{i \in I} \sum_{\sigma \in \psi'_i,0} p_i p_{i,\sigma} \frac{1}{|L(\sigma) \cap R_i|} \sum_{\rho \in L(\sigma) \cap R_i} Y(\rho, \eta(\rho))
\leq q_{k-1} \sum_{i \in I} p_i + \sum_{i \in I} \frac{1}{10 \cdot 2^g} q_{k-1} + \sum_{i \in I} \sum_{\sigma \in \psi'_i,0} p_i p_{i,\sigma} \frac{1}{10} q_{k-1}
\leq q_{k-1} + \frac{1}{10} q_{k-1} + \frac{1}{10} q_{k-1}
\leq 2q_{k-1}.
\]

We have proved (2). This completes the inductive step in the proof of the Main Lemma.

4 Proof of Theorem 1

Let \( n \geq N_0 \), where \( N_0 = 10^8 \). Suppose that \( A \in A_n \) has capacity \( S \) and time \( T \). Define \( t, k_0, m_k, q_k \) as in Section 2.3. Clearly,

\[
T \geq n - 1, \quad (7)
\]

and

\[
S \geq \lg(n - 1). \quad (8)
\]

Also, \( k_0 \leq \ln n / \ln t \leq (\ln n)^{1/2} \), and hence

\[
t^{2^{4k_0}} \leq e^{(\ln n)^{1/2} + k_0 \cdot 4 \ln 2}
\leq n^{\epsilon(n)} \quad (9)
\]

It follows that \( m_{k_0} = 2^{4k_0} + 16 t S \leq 2^{16} n^{\epsilon(n)} S \). If \( m_{k_0} \geq n/8 \), then \( S = \Omega(n^{1-\epsilon(n)}) \); hence from (7) we have \( TS = \Omega(n^{2-\epsilon(n)}) \), which proves the theorem. Thus, we can assume that

\[
m_{k_0} < \frac{n}{8}. \quad (10)
\]

We will prove that

\[
T \geq \frac{t^{k_0} n}{2m_{k_0}}. \quad (11)
\]

Suppose not. We will derive a contradiction. For each node \( v \), let \( F_v \) denote the branching program \( A[v, \min\{t^{k_0}, T - h_v\}] \), where \( h_v \) is the level number of \( v \). Let \( K_v \) be the set of \( \rho \in \Gamma(X) \)
such that $C_{\delta(v,\rho)}$ contains at least $m_{\rho_0}$ comparisons adjacent in $\rho$, where $\delta(v,\rho) = \xi_{F_{v,\rho}}$. By the corollary to the Main Lemma, $|K_v| \leq q_{\rho_0} \cdot n!$.

Thus,

$$|\bigcup_v K_v| \leq 2^S q_{\rho_0} \cdot n!$$

(12)

Since $i \geq 4$ and $t_{k_0} \leq n/4$, we have $q_{\rho_0} \leq t_{k_0}^2 2^{-10S} \leq 2^{2} 2^{-10S}$. Using (8), we have $2^S q_{\rho_0} \leq n^2(n-1)^{-9} < 1$. Therefore, (12) implies that there exists a $\rho \notin \bigcup_v K_v$. Let us input $\bar{x}_\rho$ to $A$. The total number of comparisons made by $A$ that are adjacent in $\rho$ is less than $[T/t_{k_0}^2] \cdot m_{k_0} \leq [n/2m_{k_0}] \cdot m_{k_0} \leq n/2 + m_{k_0}$, which by (10) is less than $n - 1$. This is a contradiction. We have proved (11).

It follows from (11) and (9) that

$$T \geq t_{k_0}^2 \frac{n}{2m_{k_0}}$$

$$= \Omega\left(\frac{t_{k_0}^{-1} n}{S}\right)$$

$$= \Omega\left(\frac{n^2 1}{S \cdot t^2 2^k_0}\right)$$

$$= \Omega\left(\frac{1}{S} n^{2-\epsilon(n)}\right)$$

This completes the proof of Theorem 1.

References


